

```
In [1]: #Import the required Libraries.
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

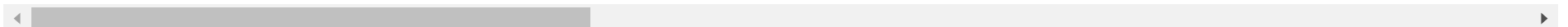
```
In [2]: #Read the data in pandas
df_1= pd.read_csv("C:\\Users\\dell-pc\\Desktop\\Oeson\\Tasks\\Task 2\\application_data.csv")
df_2= pd.read_csv("C:\\Users\\dell-pc\\Desktop\\Oeson\\Tasks\\Task 2\\previous_application.csv")
```

```
In [3]: df_app = df_1.iloc[0:30000 , 0:34]
df_app
```

```
Out[3]:
```

	SK_ID_CURR	TARGET	NAME_CONTRACT_TYPE	CODE_GENDER	FLAG_OWN_CAR	FLAG_OWN_REALTY	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_C
0	100002	1	Cash loans	M	N	Y	0	202500.0	40
1	100003	0	Cash loans	F	N	N	0	270000.0	129
2	100004	0	Revolving loans	M	Y	Y	0	67500.0	13
3	100006	0	Cash loans	F	N	Y	0	135000.0	31
4	100007	0	Cash loans	M	N	Y	0	121500.0	51
...
29995	134821	0	Cash loans	M	Y	N	2	202500.0	73
29996	134822	0	Cash loans	F	N	Y	1	225000.0	128
29997	134825	0	Cash loans	F	Y	Y	1	135000.0	26
29998	134826	0	Cash loans	F	N	Y	0	99000.0	24
29999	134827	0	Cash loans	F	N	N	0	99000.0	67

30000 rows × 34 columns

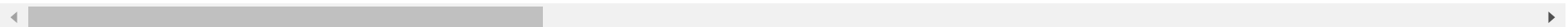


```
In [4]: df_pre = df_2.iloc[0:30000, :]
df_pre
```

```
Out[4]:
```

	SK_ID_PREV	SK_ID_CURR	NAME_CONTRACT_TYPE	AMT_ANNUITY	AMT_APPLICATION	AMT_CREDIT	AMT_DOWN_PAYMENT	AMT_GOODS_PRICE
0	2030495	271877	Consumer loans	1730.430	17145.0	17145.0	0.0	17145.0
1	2802425	108129	Cash loans	25188.615	607500.0	679671.0	NaN	607500.0
2	2523466	122040	Cash loans	15060.735	112500.0	136444.5	NaN	112500.0
3	2819243	176158	Cash loans	47041.335	450000.0	470790.0	NaN	450000.0
4	1784265	202054	Cash loans	31924.395	337500.0	404055.0	NaN	337500.0
...
29995	2006765	402412	Consumer loans	19435.455	99832.5	105102.0	0.0	99832.5
29996	2813486	161743	Consumer loans	32652.720	324000.0	324000.0	0.0	324000.0
29997	2812646	116994	Consumer loans	5656.905	40495.5	30591.0	12150.0	40495.5
29998	2127551	306088	Consumer loans	5137.560	27000.0	28336.5	0.0	27000.0
29999	2097098	302666	Consumer loans	5505.705	51975.0	46777.5	5197.5	51975.0

30000 rows × 37 columns



```
In [5]: df_app.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30000 entries, 0 to 29999
Data columns (total 34 columns):
#   Column              Non-Null Count  Dtype
---  -
0   SK_ID_CURR          30000 non-null  int64
1   TARGET              30000 non-null  int64
2   NAME_CONTRACT_TYPE  30000 non-null  object
3   CODE_GENDER         30000 non-null  object
4   FLAG_OWN_CAR        30000 non-null  object
5   FLAG_OWN_REALTY     30000 non-null  object
6   CNT_CHILDREN        30000 non-null  int64
```

```

7  AMT_INCOME_TOTAL      30000 non-null float64
8  AMT_CREDIT            30000 non-null float64
9  AMT_ANNUITY           30000 non-null float64
10 AMT_GOODS_PRICE       29976 non-null float64
11 NAME_TYPE_SUITE       29875 non-null object
12 NAME_INCOME_TYPE      30000 non-null object
13 NAME_EDUCATION_TYPE   30000 non-null object
14 NAME_FAMILY_STATUS    30000 non-null object
15 NAME_HOUSING_TYPE     30000 non-null object
16 REGION_POPULATION_RELATIVE 30000 non-null float64
17 DAYS_BIRTH            30000 non-null int64
18 DAYS_EMPLOYED         30000 non-null int64
19 DAYS_REGISTRATION     30000 non-null float64
20 DAYS_ID_PUBLISH       30000 non-null int64
21 OWN_CAR_AGE           10221 non-null float64
22 FLAG_MOBIL            30000 non-null int64
23 FLAG_EMP_PHONE        30000 non-null int64
24 FLAG_WORK_PHONE       30000 non-null int64
25 FLAG_CONT_MOBILE      30000 non-null int64
26 FLAG_PHONE            30000 non-null int64
27 FLAG_EMAIL            30000 non-null int64
28 OCCUPATION_TYPE       20554 non-null object
29 CNT_FAM_MEMBERS       30000 non-null float64
30 REGION_RATING_CLIENT  30000 non-null int64
31 REGION_RATING_CLIENT_W_CITY 30000 non-null int64
32 WEEKDAY_APPR_PROCESS_START 30000 non-null object
33 HOUR_APPR_PROCESS_START 30000 non-null int64

```

dtypes: float64(8), int64(15), object(11)

memory usage: 7.8+ MB

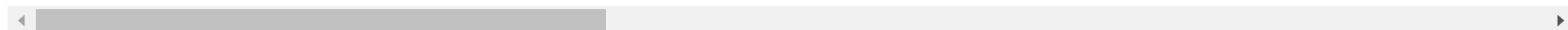
In [6]: `df_app.describe()`

Out[6]:

	SK_ID_CURR	TARGET	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE	REGION_POPULATION_RELAT
count	30000.000000	30000.000000	30000.000000	3.000000e+04	3.000000e+04	30000.000000	2.997600e+04	30000.000
mean	117466.847300	0.080100	0.416233	1.723455e+05	6.011785e+05	27177.089250	5.404012e+05	0.020
std	10065.688429	0.271452	0.722542	6.818149e+05	4.028563e+05	14674.061484	3.703965e+05	0.013
min	100002.000000	0.000000	0.000000	2.565000e+04	4.500000e+04	2052.000000	4.500000e+04	0.000
25%	108745.750000	0.000000	0.000000	1.125000e+05	2.700000e+05	16456.500000	2.385000e+05	0.010

	SK_ID_CURR	TARGET	CNT_CHILDREN	AMT_INCOME_TOTAL	AMT_CREDIT	AMT_ANNUITY	AMT_GOODS_PRICE	REGION_POPULATION_RELAT
50%	117513.500000	0.000000	0.000000	1.449000e+05	5.178555e+05	24984.000000	4.500000e+05	0.018
75%	126182.250000	0.000000	1.000000	2.025000e+05	8.100000e+05	34749.000000	6.795000e+05	0.028
max	134827.000000	1.000000	9.000000	1.170000e+08	4.050000e+06	258025.500000	4.050000e+06	0.072

8 rows × 23 columns



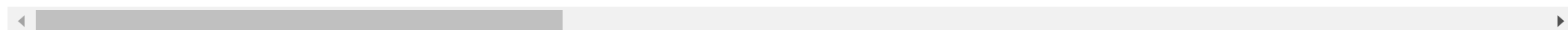
In [7]:

df_pre.describe()

Out[7]:

	SK_ID_PREV	SK_ID_CURR	AMT_ANNUITY	AMT_APPLICATION	AMT_CREDIT	AMT_DOWN_PAYMENT	AMT_GOODS_PRICE	HOUR_APPR_PROCESS
count	3.000000e+04	30000.000000	23738.000000	3.000000e+04	3.000000e+04	15137.000000	2.366900e+04	30000.
mean	1.920259e+06	278804.124800	15418.029931	1.685651e+05	1.881158e+05	6585.395081	2.136986e+05	12.
std	5.354031e+05	102755.009784	14411.480469	2.816287e+05	3.086342e+05	16052.128105	3.015334e+05	3.
min	1.000009e+06	100007.000000	0.000000	0.000000e+00	0.000000e+00	0.000000	0.000000e+00	0.
25%	1.455869e+06	189589.000000	6122.126250	2.245500e+04	2.648588e+04	0.000000	4.945500e+04	10.
50%	1.917674e+06	278329.500000	10806.390000	7.195500e+04	7.908525e+04	1575.000000	1.034550e+05	12.
75%	2.385551e+06	368501.750000	19584.011250	1.800000e+05	1.952348e+05	8091.000000	2.250000e+05	15.
max	2.845367e+06	456254.000000	210115.485000	3.150000e+06	4.104351e+06	509850.000000	3.150000e+06	23.

8 rows × 21 columns



In [8]:

df_pre.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30000 entries, 0 to 29999
Data columns (total 37 columns):
#   Column                Non-Null Count  Dtype
---  -
0   SK_ID_PREV            30000 non-null  int64
```

1	SK_ID_CURR	30000	non-null	int64
2	NAME_CONTRACT_TYPE	30000	non-null	object
3	AMT_ANNUITY	23738	non-null	float64
4	AMT_APPLICATION	30000	non-null	float64
5	AMT_CREDIT	30000	non-null	float64
6	AMT_DOWN_PAYMENT	15137	non-null	float64
7	AMT_GOODS_PRICE	23669	non-null	float64
8	WEEKDAY_APPR_PROCESS_START	30000	non-null	object
9	HOUR_APPR_PROCESS_START	30000	non-null	int64
10	FLAG_LAST_APPL_PER_CONTRACT	30000	non-null	object
11	NFLAG_LAST_APPL_IN_DAY	30000	non-null	int64
12	RATE_DOWN_PAYMENT	15137	non-null	float64
13	RATE_INTEREST_PRIMARY	97	non-null	float64
14	RATE_INTEREST_PRIVILEGED	97	non-null	float64
15	NAME_CASH_LOAN_PURPOSE	30000	non-null	object
16	NAME_CONTRACT_STATUS	30000	non-null	object
17	DAYS_DECISION	30000	non-null	int64
18	NAME_PAYMENT_TYPE	30000	non-null	object
19	CODE_REJECT_REASON	30000	non-null	object
20	NAME_TYPE_SUITE	15525	non-null	object
21	NAME_CLIENT_TYPE	30000	non-null	object
22	NAME_GOODS_CATEGORY	30000	non-null	object
23	NAME_PORTFOLIO	30000	non-null	object
24	NAME_PRODUCT_TYPE	30000	non-null	object
25	CHANNEL_TYPE	30000	non-null	object
26	SELLERPLACE_AREA	30000	non-null	int64
27	NAME_SELLER_INDUSTRY	30000	non-null	object
28	CNT_PAYMENT	23738	non-null	float64
29	NAME_YIELD_GROUP	30000	non-null	object
30	PRODUCT_COMBINATION	29994	non-null	object
31	DAYS_FIRST_DRAWING	18633	non-null	float64
32	DAYS_FIRST_DUE	18633	non-null	float64
33	DAYS_LAST_DUE_1ST_VERSION	18633	non-null	float64
34	DAYS_LAST_DUE	18633	non-null	float64
35	DAYS_TERMINATION	18633	non-null	float64
36	NFLAG_INSURED_ON_APPROVAL	18633	non-null	float64

dtypes: float64(15), int64(6), object(16)

memory usage: 8.5+ MB

In [9]:

```
print(df_app.columns.tolist())
```

```
['SK_ID_CURR', 'TARGET', 'NAME_CONTRACT_TYPE', 'CODE_GENDER', 'FLAG_OWN_CAR', 'FLAG_OWN_REALTY', 'CNT_CHILDREN', 'AMT_INCOME_TOTAL', 'AMT_CREDIT', 'AMT_ANNUITY', 'AMT_GOODS_PRICE', 'NAME_TYPE_SUITE', 'NAME_INCOME_TYPE', 'NAME_EDUCATION_TYPE', 'NAME_FAMILY_STATUS', 'NAME_REALTY_STATUS', 'NAME_SELLER_INDUSTRY', 'NAME_PORTFOLIO', 'NAME_GOODS_CATEGORY', 'NAME_CLIENT_TYPE', 'NAME_TYPE_SUITE', 'NAME_CONTRACT_STATUS', 'NAME_CASH_LOAN_PURPOSE', 'RATE_INTEREST_PRIVILEGED', 'RATE_INTEREST_PRIMARY', 'RATE_DOWN_PAYMENT', 'NFLAG_INSURED_ON_APPROVAL', 'DAYS_TERMINATION', 'DAYS_LAST_DUE', 'DAYS_LAST_DUE_1ST_VERSION', 'DAYS_FIRST_DUE', 'DAYS_FIRST_DRAWING', 'PRODUCT_COMBINATION', 'NAME_YIELD_GROUP', 'CNT_PAYMENT', 'NAME_SELLER_INDUSTRY', 'SELLERPLACE_AREA', 'CHANNEL_TYPE', 'NAME_PRODUCT_TYPE', 'NAME_PORTFOLIO', 'NAME_GOODS_CATEGORY', 'NAME_CLIENT_TYPE', 'NAME_TYPE_SUITE', 'CODE_REJECT_REASON', 'NAME_PAYMENT_TYPE', 'DAYS_DECISION', 'NAME_CONTRACT_STATUS', 'NAME_CASH_LOAN_PURPOSE', 'RATE_INTEREST_PRIVILEGED', 'RATE_INTEREST_PRIMARY', 'RATE_DOWN_PAYMENT', 'NFLAG_LAST_APPL_IN_DAY', 'FLAG_LAST_APPL_PER_CONTRACT', 'HOUR_APPR_PROCESS_START', 'WEEKDAY_APPR_PROCESS_START', 'AMT_GOODS_PRICE', 'AMT_DOWN_PAYMENT', 'AMT_CREDIT', 'AMT_APPLICATION', 'AMT_ANNUITY', 'NAME_CONTRACT_TYPE', 'SK_ID_CURR']
```

```
TUS', 'NAME_HOUSING_TYPE', 'REGION_POPULATION_RELATIVE', 'DAYS_BIRTH', 'DAYS_EMPLOYED', 'DAYS_REGISTRATION', 'DAYS_ID_PUBLISH', 'OWN_CAR_AGE', 'FLAG_MOBIL', 'FLAG_EMP_PHONE', 'FLAG_WORK_PHONE', 'FLAG_CONT_MOBILE', 'FLAG_PHONE', 'FLAG_EMAIL', 'OCCUPATION_TYPE', 'CNT_FAM_MEMBERS', 'REGION_RATING_CLIENT', 'REGION_RATING_CLIENT_W_CITY', 'WEEKDAY_APPR_PROCESS_START', 'HOUR_APPR_PROCESS_START']
```

```
In [10]: pd.options.display.max_rows = None
```

```
In [11]: print(df_app.isnull().sum())
```

```
SK_ID_CURR      0
TARGET          0
NAME_CONTRACT_TYPE  0
CODE_GENDER     0
FLAG_OWN_CAR    0
FLAG_OWN_REALTY  0
CNT_CHILDREN    0
AMT_INCOME_TOTAL  0
AMT_CREDIT      0
AMT_ANNUITY     0
AMT_GOODS_PRICE 24
NAME_TYPE_SUITE 125
NAME_INCOME_TYPE 0
NAME_EDUCATION_TYPE 0
NAME_FAMILY_STATUS 0
NAME_HOUSING_TYPE 0
REGION_POPULATION_RELATIVE 0
DAYS_BIRTH      0
DAYS_EMPLOYED   0
DAYS_REGISTRATION 0
DAYS_ID_PUBLISH 0
OWN_CAR_AGE     19779
FLAG_MOBIL      0
FLAG_EMP_PHONE  0
FLAG_WORK_PHONE 0
FLAG_CONT_MOBILE 0
FLAG_PHONE      0
FLAG_EMAIL      0
OCCUPATION_TYPE 9446
CNT_FAM_MEMBERS 0
REGION_RATING_CLIENT 0
REGION_RATING_CLIENT_W_CITY 0
WEEKDAY_APPR_PROCESS_START 0
```

```

    HOUR_APPR_PROCESS_START      0
    dtype: int64

```

In [12]:

```
print(df_pre.isnull().sum())
```

```

SK_ID_PREV      0
SK_ID_CURR      0
NAME_CONTRACT_TYPE      0
AMT_ANNUITY      6262
AMT_APPLICATION      0
AMT_CREDIT      0
AMT_DOWN_PAYMENT      14863
AMT_GOODS_PRICE      6331
WEEKDAY_APPR_PROCESS_START      0
HOUR_APPR_PROCESS_START      0
FLAG_LAST_APPL_PER_CONTRACT      0
NFLAG_LAST_APPL_IN_DAY      0
RATE_DOWN_PAYMENT      14863
RATE_INTEREST_PRIMARY      29903
RATE_INTEREST_PRIVILEGED      29903
NAME_CASH_LOAN_PURPOSE      0
NAME_CONTRACT_STATUS      0
DAYS_DECISION      0
NAME_PAYMENT_TYPE      0
CODE_REJECT_REASON      0
NAME_TYPE_SUITE      14475
NAME_CLIENT_TYPE      0
NAME_GOODS_CATEGORY      0
NAME_PORTFOLIO      0
NAME_PRODUCT_TYPE      0
CHANNEL_TYPE      0
SELLERPLACE_AREA      0
NAME_SELLER_INDUSTRY      0
CNT_PAYMENT      6262
NAME_YIELD_GROUP      0
PRODUCT_COMBINATION      6
DAYS_FIRST_DRAWING      11367
DAYS_FIRST_DUE      11367
DAYS_LAST_DUE_1ST_VERSION      11367
DAYS_LAST_DUE      11367
DAYS_TERMINATION      11367
NFLAG_INSURED_ON_APPROVAL      11367
dtype: int64

```

In [13]:

```
df_app = df_app.ffill(axis = 0)
print(df_app.isnull().sum())
```

```
SK_ID_CURR      0
TARGET          0
NAME_CONTRACT_TYPE  0
CODE_GENDER     0
FLAG_OWN_CAR    0
FLAG_OWN_REALTY  0
CNT_CHILDREN    0
AMT_INCOME_TOTAL  0
AMT_CREDIT      0
AMT_ANNUITY     0
AMT_GOODS_PRICE  0
NAME_TYPE_SUITE  0
NAME_INCOME_TYPE  0
NAME_EDUCATION_TYPE  0
NAME_FAMILY_STATUS  0
NAME_HOUSING_TYPE  0
REGION_POPULATION_RELATIVE  0
DAYS_BIRTH      0
DAYS_EMPLOYED    0
DAYS_REGISTRATION  0
DAYS_ID_PUBLISH  0
OWN_CAR_AGE      2
FLAG_MOBIL      0
FLAG_EMP_PHONE   0
FLAG_WORK_PHONE  0
FLAG_CONT_MOBILE  0
FLAG_PHONE       0
FLAG_EMAIL       0
OCCUPATION_TYPE  0
CNT_FAM_MEMBERS  0
REGION_RATING_CLIENT  0
REGION_RATING_CLIENT_W_CITY  0
WEEKDAY_APPR_PROCESS_START  0
HOUR_APPR_PROCESS_START  0
dtype: int64
```

In [14]:

```
df_pre = df_pre.ffill(axis = 0)
print(df_pre.isnull().sum())
```


SK_ID_PREV	0
SK_ID_CURR	0
NAME_CONTRACT_TYPE	0
AMT_ANNUITY	0
AMT_APPLICATION	0
AMT_CREDIT	0
AMT_DOWN_PAYMENT	0
AMT_GOODS_PRICE	0
WEEKDAY_APPR_PROCESS_START	0
HOUR_APPR_PROCESS_START	0
FLAG_LAST_APPL_PER_CONTRACT	0
NFLAG_LAST_APPL_IN_DAY	0
RATE_DOWN_PAYMENT	0
RATE_INTEREST_PRIMARY	0
RATE_INTEREST_PRIVILEGED	0
NAME_CASH_LOAN_PURPOSE	0
NAME_CONTRACT_STATUS	0
DAYS_DECISION	0
NAME_PAYMENT_TYPE	0
CODE_REJECT_REASON	0
NAME_TYPE_SUITE	1
NAME_CLIENT_TYPE	0
NAME_GOODS_CATEGORY	0
NAME_PORTFOLIO	0
NAME_PRODUCT_TYPE	0
CHANNEL_TYPE	0
SELLERPLACE_AREA	0
NAME_SELLER_INDUSTRY	0
CNT_PAYMENT	0
NAME_YIELD_GROUP	0
PRODUCT_COMBINATION	0
DAYS_FIRST_DRAWING	0
DAYS_FIRST_DUE	0
DAYS_LAST_DUE_1ST_VERSION	0
DAYS_LAST_DUE	0
DAYS_TERMINATION	0
NFLAG_INSURED_ON_APPROVAL	0

dtype: int64

APPLICANT DATA SET

DATA ANALYSIS

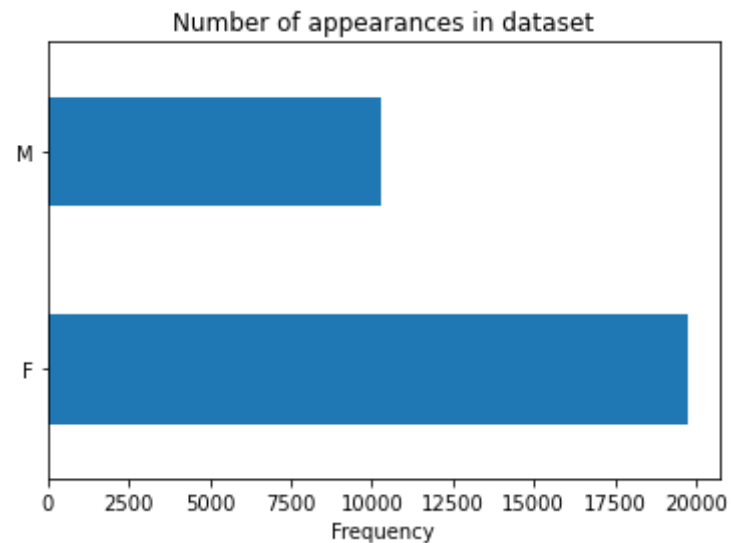
```
In [15]: df_app['CODE_GENDER'] = df_app['CODE_GENDER'].astype('string')
```

```
In [16]: df_app['CODE_GENDER'].dtypes
```

```
Out[16]: string[python]
```

```
In [17]: df_app.CODE_GENDER.value_counts().plot(kind='barh')
plt.title('Number of appearances in dataset')
plt.xlabel('Frequency')
```

```
Out[17]: Text(0.5, 0, 'Frequency')
```



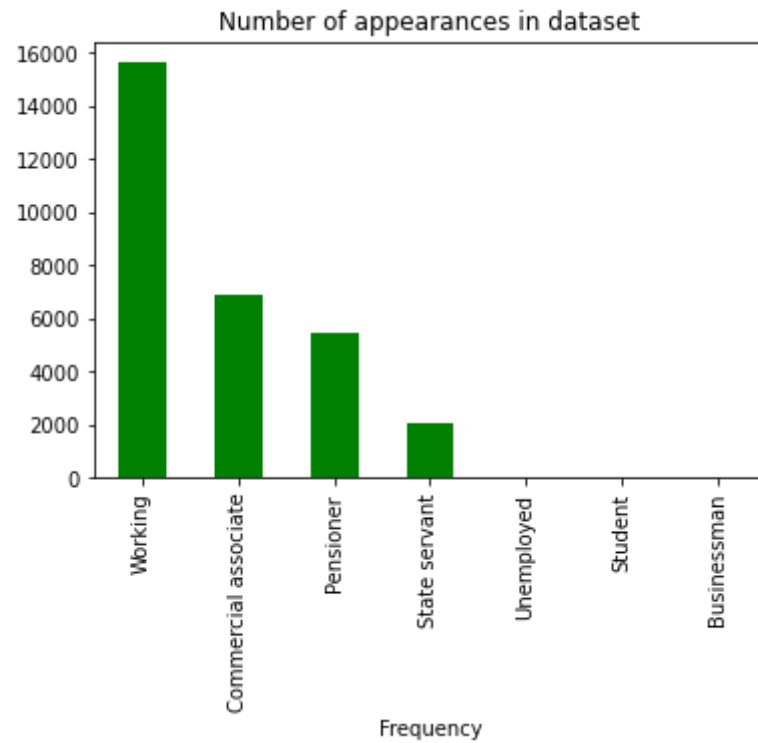
```
In [18]: df_app['AMT_INCOME_TOTAL'] = df_app['AMT_INCOME_TOTAL'].astype(float)
```

```
In [19]: df_app['AMT_INCOME_TOTAL'].dtypes
```

```
Out[19]: dtype('float64')
```

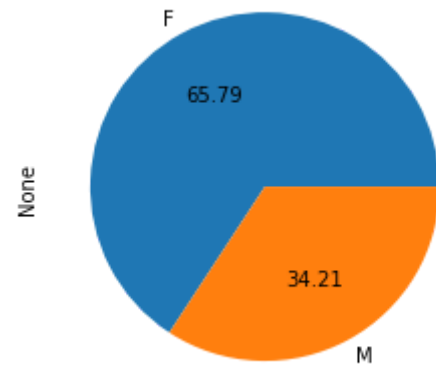
```
In [20]: df_app.NAME_INCOME_TYPE.value_counts().plot(kind='bar', color = "green")
plt.title('Number of appearances in dataset')
plt.xlabel('Frequency')
```

```
Out[20]: Text(0.5, 0, 'Frequency')
```



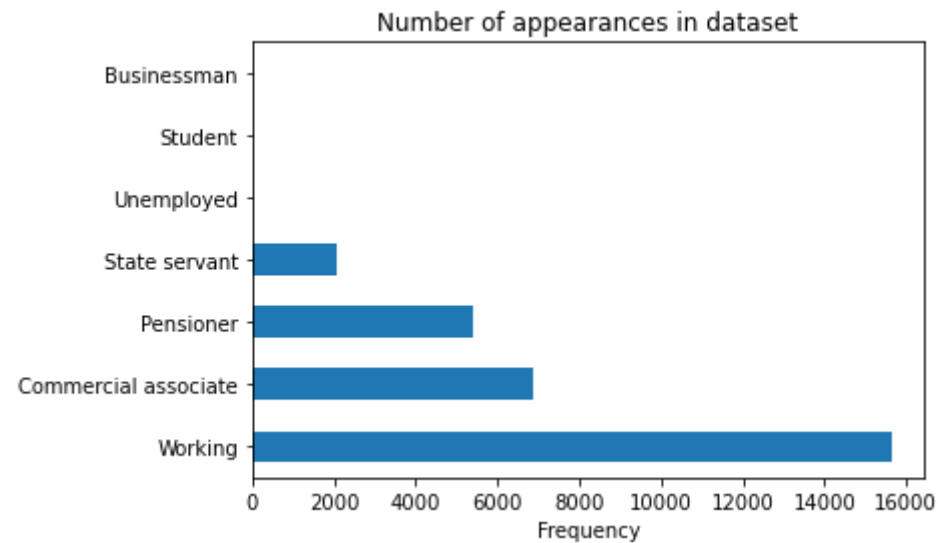
```
In [21]: df_app.groupby('CODE_GENDER').size().plot(kind='pie', autopct='%.2f')
```

```
Out[21]: <AxesSubplot:ylabel='None'>
```

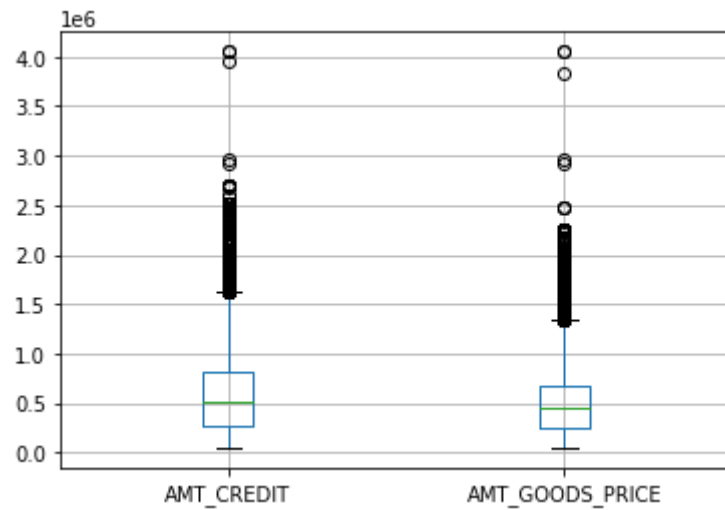


```
In [22]: df_app.NAME_INCOME_TYPE.value_counts().plot(kind='barh')
plt.title('Number of appearances in dataset')
plt.xlabel('Frequency')
```

```
Out[22]: Text(0.5, 0, 'Frequency')
```



```
In [23]: boxplot = df_app.boxplot(column=['AMT_CREDIT', 'AMT_GOODS_PRICE'])
```



```
In [24]: df_app.value_counts(["NAME_FAMILY_STATUS", "TARGET"])
```

```
Out[24]: NAME_FAMILY_STATUS  TARGET
Married                0      17911
Single / not married   0      3879
Civil marriage         0      2601
Separated              0      1725
Widow                  0      1481
Married                1      1427
Single / not married   1       441
Civil marriage         1       285
Separated              1       164
Widow                  1        86
dtype: int64
```

```
In [25]: pd.crosstab(df_app.CODE_GENDER, df_app.TARGET)
```

```
Out[25]:
```

	TARGET	0	1
CODE_GENDER			
F	18369	1368	
M	9228	1035	

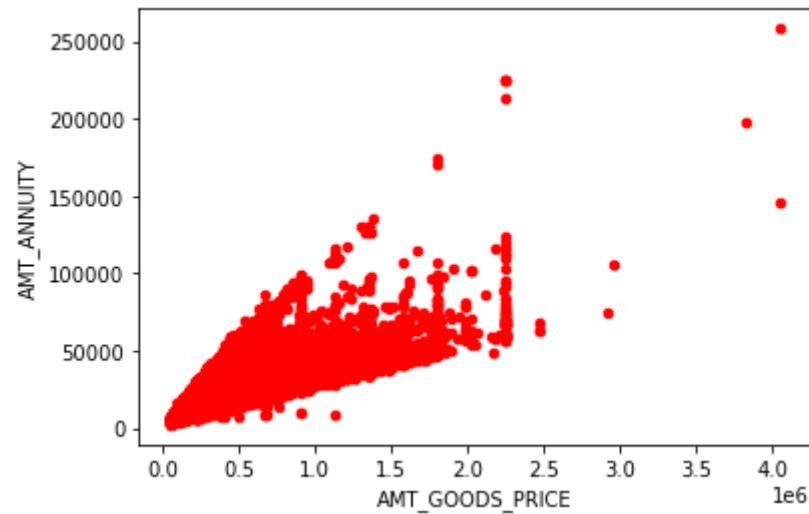
```
In [26]: df_app.value_counts(["NAME_INCOME_TYPE", "TARGET"])
```

```
Out[26]: NAME_INCOME_TYPE    TARGET
Working                0      14172
Commercial associate  0       6349
Pensioner             0       5113
State servant         0       1957
Working               1       1470
Commercial associate  1        513
Pensioner            1       308
State servant        1       111
Businessman          0         2
Student              0         2
Unemployed           0         2
                    1         1
dtype: int64
```

```
In [27]: scatter = df_app.plot.scatter(x='AMT_GOODS_PRICE',y='AMT_ANNUITY', c='Red')
core = df_app[['AMT_GOODS_PRICE', 'AMT_ANNUITY']].corr()
core
```

```
Out[27]:
```

	AMT_GOODS_PRICE	AMT_ANNUITY
AMT_GOODS_PRICE	1.000000	0.778179
AMT_ANNUITY	0.778179	1.000000



```
In [28]: corr = df_app[['AMT_INCOME_TOTAL', 'AMT_GOODS_PRICE', 'AMT_CREDIT']].corr()
corr
```

```
Out[28]:
```

	AMT_INCOME_TOTAL	AMT_GOODS_PRICE	AMT_CREDIT
AMT_INCOME_TOTAL	1.000000	0.054498	0.054326
AMT_GOODS_PRICE	0.054498	1.000000	0.986357
AMT_CREDIT	0.054326	0.986357	1.000000

```
In [29]: corr.style.background_gradient(cmap='coolwarm')
```

```
Out[29]:
```

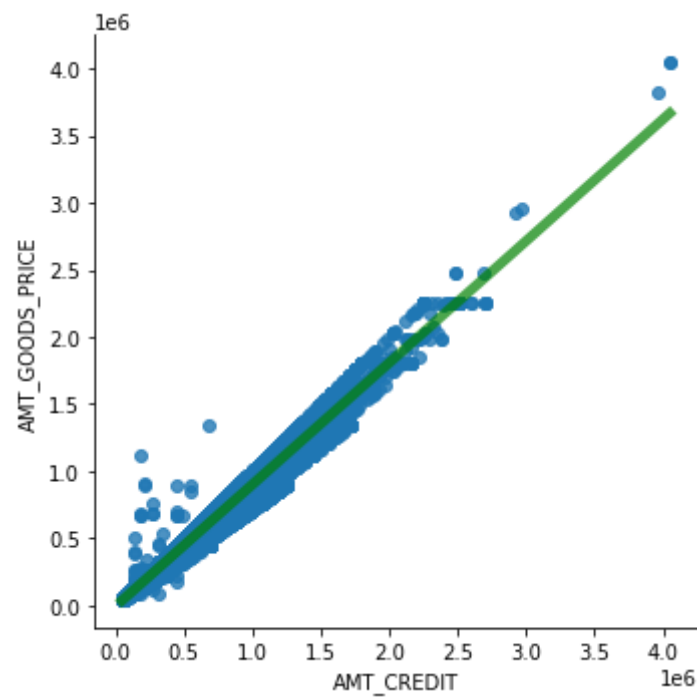
	AMT_INCOME_TOTAL	AMT_GOODS_PRICE	AMT_CREDIT
AMT_INCOME_TOTAL	1.000000	0.054498	0.054326
AMT_GOODS_PRICE	0.054498	1.000000	0.986357
AMT_CREDIT	0.054326	0.986357	1.000000

```
In [30]: sns.lmplot(x='AMT_CREDIT', y='AMT_GOODS_PRICE', data=df_app, ci=None, line_kws={"color": "g", "alpha": 0.7, "lw": 5})
corr = df_app[['AMT_CREDIT', 'AMT_GOODS_PRICE']].corr()
```

```
cort
```

```
Out[30]:
```

	AMT_CREDIT	AMT_GOODS_PRICE
AMT_CREDIT	1.000000	0.986357
AMT_GOODS_PRICE	0.986357	1.000000

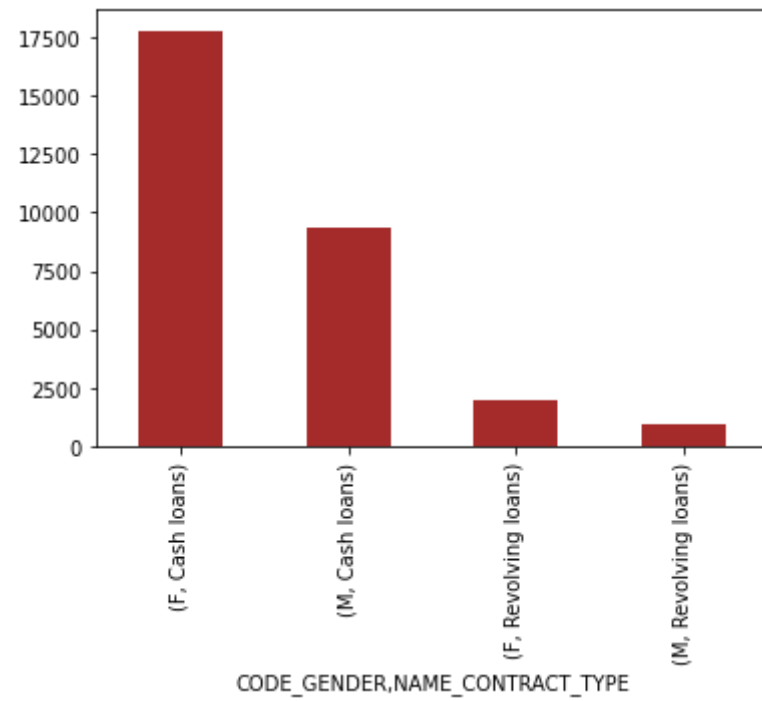


```
In [31]:
```

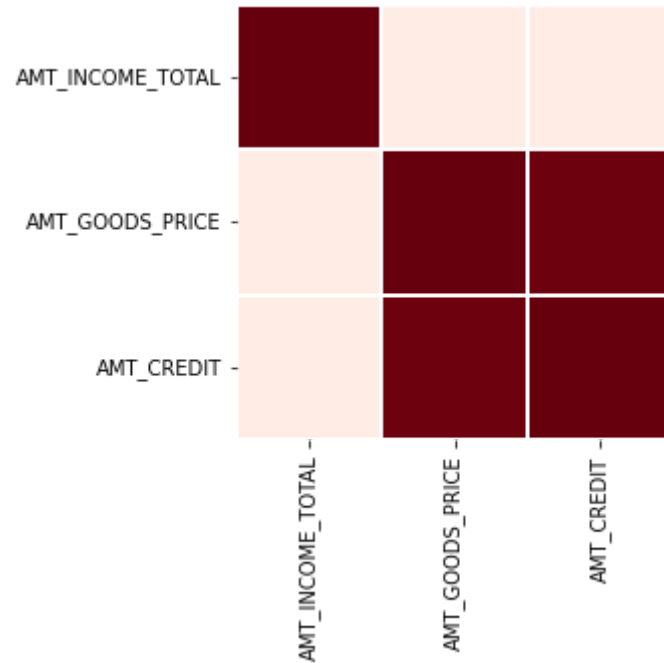
```
df_app.value_counts(["CODE_GENDER", "NAME_CONTRACT_TYPE"]).plot(kind='bar', color= 'brown')
```

```
Out[31]:
```

```
<AxesSubplot:xlabel='CODE_GENDER,NAME_CONTRACT_TYPE'>
```

```
In [32]: heatmap = sns.heatmap(corr, cbar=0, linewidths=2, vmax=1, vmin=0, square=True, cmap='Reds')  
plt.show()
```



PREVIOUS DATA

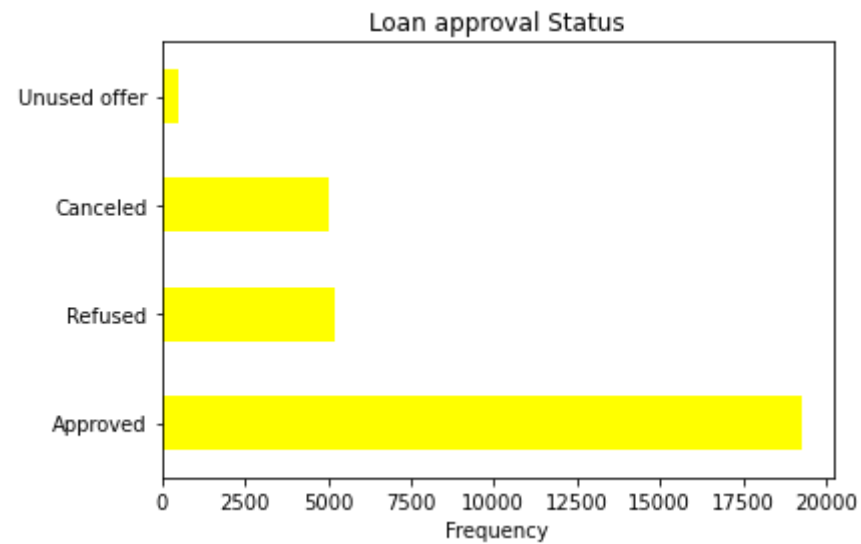
DATA ANALYSIS

```
In [33]: print(df_pre.columns.tolist())
```

```
['SK_ID_PREV', 'SK_ID_CURR', 'NAME_CONTRACT_TYPE', 'AMT_ANNUITY', 'AMT_APPLICATION', 'AMT_CREDIT', 'AMT_DOWN_PAYMENT', 'AMT_GOODS_PRICE', 'WEEKDAY_APPR_PROCESS_START', 'HOUR_APPR_PROCESS_START', 'FLAG_LAST_APPL_PER_CONTRACT', 'NFLAG_LAST_APPL_IN_DAY', 'RATE_DOWN_PAYMENT', 'RATE_INTEREST_PRIMARY', 'RATE_INTEREST_PRIVILEGED', 'NAME_CASH_LOAN_PURPOSE', 'NAME_CONTRACT_STATUS', 'DAYS_DECISION', 'NAME_PAYMENT_TYPE', 'CODE_REJECT_REASON', 'NAME_TYPE_SUITE', 'NAME_CLIENT_TYPE', 'NAME_GOODS_CATEGORY', 'NAME_PORTFOLIO', 'NAME_PRODUCT_TYPE', 'CHANNEL_TYPE', 'SELLERPLACE_AREA', 'NAME_SELLER_INDUSTRY', 'CNT_PAYMENT', 'NAME_YIELD_GROUP', 'PRODUCT_COMBINATION', 'DAYS_FIRST_DRAWING', 'DAYS_FIRST_DUE', 'DAYS_LAST_DUE_1ST_VERSION', 'DAYS_LAST_DUE', 'DAYS_TERMINATION', 'NFLAG_INSURED_ON_APPROVAL']
```

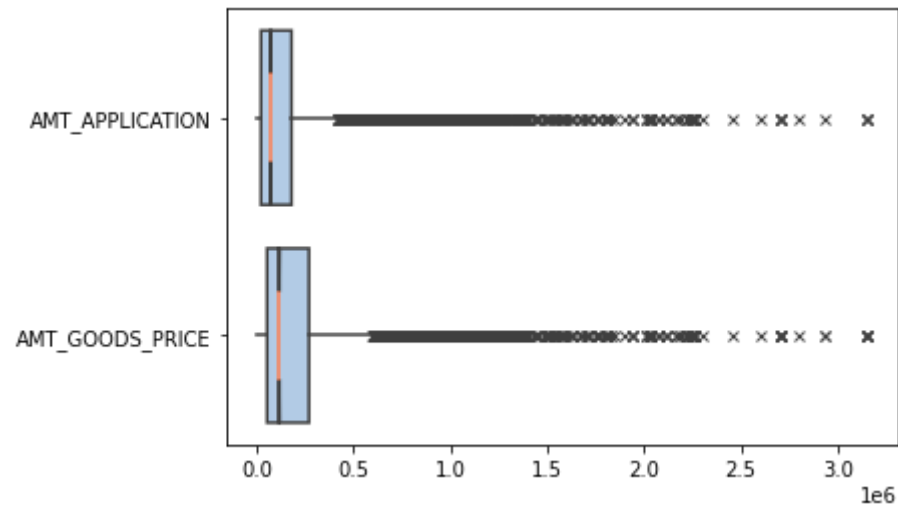
```
In [34]: df_pre.NAME_CONTRACT_STATUS.value_counts().plot(kind='barh', color = 'Yellow')
plt.title('Loan approval Status')
plt.xlabel('Frequency')
```

Out[34]: Text(0.5, 0, 'Frequency')



```
In [35]: a = sns.boxplot(data = df_pre[["AMT_APPLICATION", "AMT_GOODS_PRICE"]], orient="h", notch=True, showcaps=False,
    flierprops={"marker": "x"},
    boxprops={"facecolor": (.4, .6, .8, .5)},
    medianprops={"color": "coral"},)
a
```

Out[35]: <AxesSubplot:>



```
In [36]: pd.crosstab(df_pre.NAME_CONTRACT_TYPE,df_pre.NAME_CONTRACT_STATUS)
```

```
Out[36]: NAME_CONTRACT_STATUS  Approved  Canceled  Refused  Unused offer
```

NAME_CONTRACT_TYPE				
Cash loans	5254	4220	2837	9
Consumer loans	12330	35	1479	517
Revolving loans	1670	773	870	0
XNA	0	6	0	0

```
In [37]: corr2 = df_pre[['AMT_ANNUITY', 'AMT_APPLICATION', 'AMT_CREDIT']].corr()
corr2
```

```
Out[37]:
```

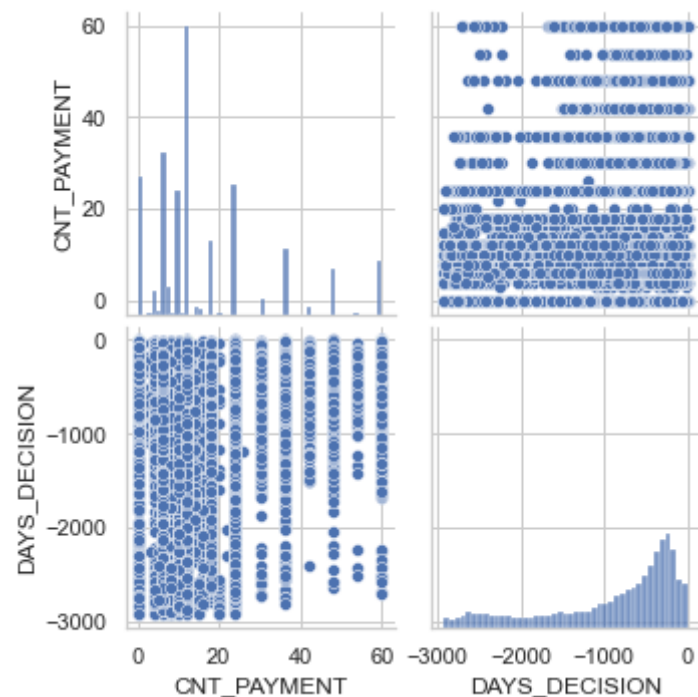
	AMT_ANNUITY	AMT_APPLICATION	AMT_CREDIT
AMT_ANNUITY	1.000000	0.619339	0.622485
AMT_APPLICATION	0.619339	1.000000	0.976210
AMT_CREDIT	0.622485	0.976210	1.000000

```
In [38]: corr2.style.background_gradient(cmap='Oranges')
```

```
Out[38]:
```

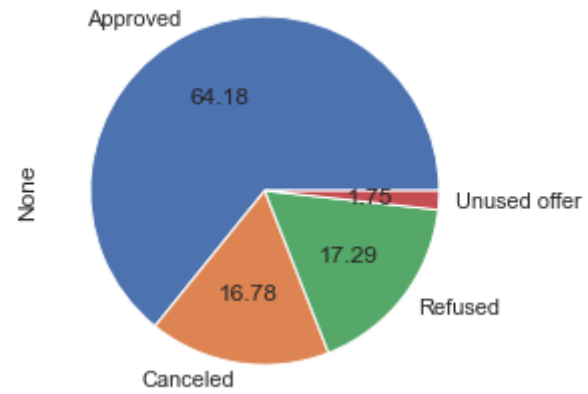
	AMT_ANNUIITY	AMT_APPLICATION	AMT_CREDIT
AMT_ANNUIITY	1.000000	0.619339	0.622485
AMT_APPLICATION	0.619339	1.000000	0.976210
AMT_CREDIT	0.622485	0.976210	1.000000

```
In [39]: sns.set_theme(style="whitegrid", palette="deep")
sns.pairplot(data= df_pre, vars= ["CNT_PAYMENT", "DAYS_DECISION"])
plt.show()
```



```
In [40]: df_pre.groupby('NAME_CONTRACT_STATUS').size().plot(kind='pie', autopct='%.2f')
```

```
Out[40]: <AxesSubplot:ylabel='None'>
```



In []: